

COMP 630: Databases

Fall 2020

Professor: John Greiner

Contact Information: greiner@rice.edu, Duncan Hall 3093

TA: Zhengtong Liu

Contact Information: zl92@rice.edu

Course Materials

All required course materials will be available via the Canvas web site.

Note that I do not post the PowerPoint slides that correspond to many of the lecture videos. I generally find that this reduces student attention to the lecture contents. As a compromise, I post “Notes Handouts”, which are summaries of these slides with space for taking notes. I do post the Jupyter Notebooks that correspond to other lecture videos because these contain many programming examples that you may want to experiment with.

This course does not use a textbook. If you would like a textbook for supplemental reading, there are a lot of good ones available. “Fundamentals of Database Systems” by Elmasri and Navathe is a good text on the broad subject area. Joe Celko’s series of “SQL for Smarties” are good for a deeper look specifically at SQL.

Software Requirements

All the course software will be available via the web-based programming environment Vocareum. All programming-based assignments will be submitted via Vocareum. We will use the database systems PostgreSQL, MongoDB, and Neo4J, plus the tool Jupyter Notebook, plus libraries to link the database systems to Jupyter.

Some assignments require a small amount of programming in an application language (either Python or Java). We assume students are at least somewhat familiar with programming in such

a language. Most programming will be done in database languages that will be introduced in the course.

Downloading the course software onto your own computer is not necessary, but optional. All of the software is available for free.

Course Overview

The course has five learning objectives, i.e., high-level goals:

- (Big picture) Understand the trade-offs of relational and non-relational databases
- (Queries) Manage data and understand the costs of doing so
- (Design) Design complex databases and understand design trade-offs
- (Real-world data) Curate and merge data from real-world sources
- (Communication) Explain concepts and implementation and design decisions

Prerequisites

There are no specific course prerequisites, but there are general concepts that it would be best to have prior to taking this course. In addition to high school math, the following material will be used.

Discrete Math

- “Big-O”: Familiar with the concept of $O()$, but not necessarily the formal definitions of asymptotes. Familiar with the terms *logarithmic*, *linear*, *quadratic*.
- First-order logic: Can interpret and create logical expressions with quantification. Can write short proofs.
- Sets: Can use set operations and notation.
- Functions and relations: Familiar with the mathematical definitions.

Programming

- Data structures: Familiar with the concepts, but not necessarily the implementations, of arrays, linked lists, trees, balanced search trees, graphs, mappings.
- Algorithms: Familiar with the concepts, but not necessarily the implementations, of linear search, binary search, sorting in $O(n \log n)$ time. Can understand an algorithm presented in pseudo-code.
- Skills: Can write 25-line programs in an application language, preferably Python or Java, and be able to call provided APIs. Has had some experience using a command-line interface. Can use and understand a programming reference manual.

Course Requirements

Activity	Deadline	Grade weight	Time estimate
Videos + in-video questions		0%	2 hrs/wk
Quizzes (27)	Wed. night	10%	1 hr/wk
Exercise sets (27)	Wed. night	14%	1 hr/wk
Feedback survey (14)	Wed. night	2%	0.25 hrs/wk
Live problem session (14)	Thursday	0%	1 hr/wk
Assignments (6)	Sun. night	40%	8 hrs each
Exams (2)		34%	3 hrs each

The time required for each activity, especially the assignments, can vary tremendously by student. The times listed in the above table are only estimates.

Your activities in this class are organized around the following schedule. However, the material is released early so that you will have lots of flexibility to do the activities prior to the deadlines.

Monday - Wednesday: Each week will have two sets of material, or topics, each consisting of

- A set of video lectures
 - Some videos contain in-video questions. These are not graded, but the course staff will see and review your answers.
 - Some videos are followed by supplemental text to further explain the content. Be sure to scroll down below each video to look for such text.
- A “notes handout” -- This summarizes some of the important points from the videos, provides you a place to take notes, and repeats the information needed to complete the in-video questions.
- A quiz on the video content

- A set of exercises going into greater depth than the quiz -- While these count towards your grade, I strongly encourage you to think of them as part of the learning process in preparation for the assignments.

Each week will also have a short anonymous feedback survey. You will provide feedback to the instructor and, in doing so, also reflect on the week's material.

The quizzes, exercises, and feedback survey are all due Wednesday night.

Thursday: We will host a live interactive session with group problem-solving and Q&A discussion. The primary purpose is to resolve any confusion students have on the material. Participation is strongly encouraged, but optional.

Sunday: Roughly every two weeks, an assignment will be due. These will be due Sunday night.

Other: There will also be a midterm exam at the half-way point, plus a final exam after the last week.

Attendance Policy

Attendance is not required, but strongly encouraged for all live class sessions. Your grade will not be penalized if you don't attend.

However, a primary benefit of this program is the time spent during these live sessions interacting with your classmates and faculty, with much of the learning coming from these interactions. These live sessions are a major part of not only learning the content, but also the experience as a whole. As stated above, one of their primary purposes is to clarify any misconceptions that you might have. Additional content examples and details beyond the video lectures will be in the live sessions.

Contacting Staff

For non-personal questions, such as content and assignment clarifications, the best forum is Canvas Discussions. After all, if you have a question, it is likely that other students have the same question and would like to know the answer.

For personal questions about your assignment, grades, or such, the best way to contact me is via email, greiner@rice.edu. Please use a detailed subject line including "COMP 630", since I teach multiple courses.

Course Outline

The course will consist of six modules. Each module concludes with an assignment. The following summarizes the schedule and topics.

Week	Module	Topics	Assignment due
1	1	Database & relational concepts 1	
	1	Database & relational concepts 2	
2	1	SQL 1, Simple queries	
	1	SQL 2, Table and data operations	1: 9/6/2020
3	2	Data modeling overview	
	2	Entity-relationship modeling basics	
4	2	Entity-relationship modeling classes	
	2	Implementing ERDs in SQL	2 initial milestone: 9/20/2020
5	2	Normalization background	
	2	Normalization	2: 9/27/2020
6	3	SQL 3, Computation in queries	
	3	SQL 4, Multisets and quantifiers	
7	3	SQL 5, More joins & binding forms	
	4	SQL 6, Functions & triggers	3: 10/11/2020
8	4	Database/application interface	
			midterm: 10/18/2020
9	4	Transactions 1, Concurrency	
	4	Transactions 2, Recovery	
10	4	Security; Using real-world data	
	5	Data access costs 1, Storage	4: 11/1/2020
11	5	Data access costs 2, Indices	
	5	Query costs 1, Optimization	

12	5	Query costs 2, Tuning	
	6	Data warehouses, data lakes, & OLAP	5: 11/15/2020
13	6	NoSQL	
	6	MongoDB 1, Basics	
14	6	MongoDB 2, More	
	6	Neo4J	6: 12/6/2020
			Final: 12/13/2020

Assessment Deadlines and Grades

We publish course materials well before the corresponding deadlines so that students have the flexibility to work according to their own schedule. However, that flexibility ends at the deadline. Graded work will not be accepted after the posted deadline, unless there are exceptional and unforeseen circumstances.

Assignment Details

Assn.	Due Date	Points	Summary
1	9/6/2020	55	Queries in relational calculus, relational algebra, & SQL Short answer problems on basic concepts
2	9/20/2020	10	Designing a genealogical DB: entity-relationship design and table schemas
	9/27/2020	90	Short answer problems on design & normalization
3	10/11/2019	100	Complex SQL queries on the genealogical DB Short answer problems on queries
4	11/1/2019	100	Parts of a transactional application on the genealogical DB Short answer problems on DB/application integration & transactions
5	11/15/2020	100	Short answer problems on costs & efficiency
6	12/16/2020	100	OLAP, MongoDB, and Neo4J design of genealogical DB MongoDB and Neo4J queries

			Short answer problems
--	--	--	-----------------------

Canvas deadlines

Canvas has a “due by” and an “available until” deadline. The Canvas system will allow submissions until the “available until” deadline, but any submissions after the “due by” deadline will be marked late. As described, graded work will not be accepted after the posted deadline, so these two deadlines are the same.

Note that a posted deadline of 11:00pm, for example, is indeed zero seconds after 11:00pm. Canvas and the instructors will count an assignment late if it is turned in one or more seconds after the “due by” deadline, even though it may appear to be the same time. Also, the submission process takes some (small) amount of time, and your work is not submitted until the end of that process. So, be sure to **always** submit a few minutes before the deadline.

For most assessments, you will be allowed to make multiple submissions. Canvas allows you to look at your submission, and we **strongly recommend** that you check you successfully submitted what you intended. For quizzes, since they are automatically graded by computer, the **highest** score will be used. For exercises and assignments, the **last** submission will be graded.

Late submissions and extensions

As stated above, late submissions are not accepted. However, you may request an extension if there are exceptional and unforeseen circumstances. You do so by emailing the instructor with a justification and the designed extension length. Examples include emergency medical issues by immediate family or dependents, last-minute required work travel, and natural disasters. It is expected that students will have other classes, a job, families, or other responsibilities, and deadlines from such activities are **not** considered exceptional circumstances. You should request an extension as early as possible, as extension requests received after the deadline are received much less favorably.

The instructor has the flexibility to grant a class-wide extension in the event of unforeseen circumstances that affect a large portion of the class.

Grading

You will receive your assessment scores and feedback through the Canvas system. You can look either at the original assessment or go via the Canvas “Grades” feature.

You may request a regrading of your work if you think it was graded incorrectly. You do so by emailing the instructor with a specific description of which deductions you think were incorrect

and why. Examples include that the grader overlooked or misinterpreted part of your answer. A regrade request must come within 7 days of your score and feedback being posted within Canvas.

Honor Code Policy

The Honor Code is a special privilege and responsibility at Rice University. You pledged to uphold the Rice Honor Code when you matriculated at Rice, and we take the Honor Code very seriously. The work you submit for this class is expected to be the result of your own work. Attempting to take credit for someone else's work by turning it in as your own constitutes plagiarism, as defined by the Rice Honor Code.

As stated, the guiding principle is that all graded work is to be done individually. However, you are encouraged to discuss problems with the course staff and classmates. The specific limits on access to materials varies by type of assessment.

All Masters of Computer Science students are sworn to adhere to the tenets set forth in Rice University's Honor Code. All written assignments for this class submitted to the instructor must carry the signed Honor Code pledge: "On my honor, I have neither given nor received any unauthorized aid on this (examination, quiz or paper)." (with signature and/or student ID number). Failure to include this pledge will result in an ungraded and returned assignment.

Please be aware of the Honor Code addendum that is part of the Student Handbook. This will provide more specific language about how the Honor Code should be adhered to as part of the Computer Science program.

All graded work other than exams

Allowed:

- Any material provided as part of the course, including videos, notes, and solutions.
- Your course work, including graded work and any notes.
- Any website, book, or other material. However, you may use at most the equivalent of two lines of code or math per assignment (or other work), and you must reinterpret any text in your own words. You must cite any source used.
- Communicating with course staff
- Communicating with other students in the class about general approaches to problems, but not about specific solutions.

Not allowed (an incomplete list):

- Copying large parts of solutions for any source other than what is provided as part of the course or your own work
- Copying small parts of solutions *without citation* from any source other than what is provided as part of the course or your own work.

- Communicating with other students in the class about a specific solution
- Communicating with anyone outside the class. This includes doing things like posting questions to StackOverflow.
- Posting sample solutions to any public forum.

Exams

Allowed:

- Any material provided as part of the course, including videos, notes, and solutions.
- Your course work, including graded work and any notes.
- The official reference material for any application used as part of the course. The course will provide a specific list.
- Communicating with course staff

Not allowed (an incomplete list):

- Any other website, book, or other material
- Communicating with any other person about the exam
- Posting sample solutions to any public forum.

Exams will use the ProctorU service to monitor students. This requires that students use a Windows or MacOS computer for the exam.

Disability-based Accommodations

If you have a documented disability that may affect academic performance, you should: 1) make sure this documentation is on file with Disability Resource Center (Allen Center, Room 111 / adarice@rice.edu / x5841) to determine the accommodations you need; and 2) meet with me to discuss your accommodation needs.

Title IX Policy

Rice University cares about your wellbeing and safety. Rice encourages any student who has experienced an incident of harassment, pregnancy discrimination, gender discrimination, or relationship, sexual, or other forms interpersonal violence to seek support through The SAFE Office. Students should be aware when seeking support on campus that most employees, including myself, as the instructor/TA, are required by Title IX to disclose all incidents of non-consensual interpersonal behaviors to Title IX professionals on campus who can act to support that student and meet their needs. For more information, please visit safe.rice.edu or email titleixsupport@rice.edu.